

## COMMENTARY

## Open Access



# More than a response to Andrew Sampson's (2012) "Coded and uncoded error feedback: effects on error frequencies in adult Colombian EFL learners' writing": a call for replication

Barry Lee Reynolds<sup>1\*</sup> and Chian-Wen Kao<sup>2</sup>

\* Correspondence:

[barryreynolds@umac.mo](mailto:barryreynolds@umac.mo)

<sup>1</sup>Faculty of Education, University of Macau, Room 1017, FED Building (E33), Avenida da Universidade, Taipa, Macau SAR

Full list of author information is available at the end of the article

## Abstract

In an article published in *System* Vol. 40 Andrew Sampson (2012) made several claims regarding the positive effect of "coding" or "marking" of second language writing errors and how the use of said coding can have a positive effect on the number of errors appearing in L2 writers' subsequent writings. However, upon closer examination of the article's methodology, we feel such a claim regarding the use of coding in the L2 writing classroom is not justified without further research. In this commentary, through reexamination of Sampson's research, we argue that (1) correction of errors that appeared on previous drafts should not be equated with the ability to produce correct forms in future writings; (2) equality of sampling across learners' texts should have been more systematic; and (3) error types deserve a more systematic classification scheme. We further elaborate on the flaws found in the research methodology and where appropriate suggest alternatives. Finally, we conclude with some suggestions regarding coded versus uncoded feedback.

**Keywords:** L2 writing, Error correction, Coding, Feedback, Methodology

## Background

Among the traditional four skills, writing seems to be the skill that is the most difficult for learners to master and for teachers to teach. The teaching of writing is laborious because it requires some type of marking (i.e., feedback) of learners' writings and the learning of writing requires scrutiny and review of these markings. The ultimate goal of both the learners and the teachers is to have learners produce writing that effectively communicates a message. EFL writing teachers are seeking effective methods of providing learners with feedback that ultimately results with them producing future writing that is more competent than their previous writings. To accomplish this goal, many EFL writing teachers utilize a process approach to the teaching of writing that incorporates a system that indicates when learners have made grammatical errors. Many of these teachers settle on marking strategies after consultation from language teaching resource books or second language acquisition literature published in authoritative journals. Still others take TEFL courses and use

suggestions provided by their professors when they comment on their own students' writings. Thus, it is important that any methodologies or ideas that are recommended in the literature or teacher resource books are research based (Nation 2009). As teachers of EFL and TEFL, we are constantly looking for new methods to introduce to our EFL learners and pre-service EFL teachers. Andrew Sampson's (2012) "Coded and uncoded error feedback: Effects on error frequencies in adult Colombian EFL learners' writing" (*System*, Vol. 40), is one example of such an article that introduces one such marking method. The article reports a study comparing the effects of uncoded and coded corrections on Colombian EFL learners' writing. When we first ran across the article, we felt the method of coding grammar errors might be a practical approach we could introduce in EFL teacher training courses. Although we agree that it is timely that research investigating the effects of "coding" or "marking" of second language writing errors is being conducted, upon closer inspection of the article we found weaknesses in its research design and methodology that invites some further comment that could lead to more robust future research. In this paper we argue that (1) correction of errors that appeared on previous drafts should not be equated with the ability to produce correct forms in future writings; (2) equality of sampling across learners' texts should have been more systematic; and (3) error types deserve a more systematic classification scheme. Below we elaborate on these flaws in the research methodology and where appropriate suggest alternatives. Finally, we conclude with some suggestions on coded versus uncoded feedback. Having results from a more methodologically sound study would lend support to whether employing the use of coding of students' grammar errors in the EFL writing classroom would result in a return rate worth the investment by teachers.

## Discussion

### Improvement or memorization

Firstly, Sampson (2012), claims, "the experimental group were able to locate and correct their own errors slightly more successfully...than the control group... This small difference could be interpreted as suggesting coded feedback is slightly more successful at developing receptive awareness of correct forms than uncoded correction" (p. 499). Sampson (2012) assumes that after exposure to the corrected draft (whether coded or uncoded) that an L2 writer's future ability to locate and correct the errors previously marked by the teacher equals development of receptive awareness of correct forms. This is unable to be determined without further investigation. Specifically, any corrections made by an L2 writer on the original draft after viewing of the teacher's feedback on that draft may simply be due to memorization of the teacher's feedback instead of development of receptive awareness of correct forms. In other words, coded correction may have just been slightly more effective at aiding L2 writers in memorizing which errors they had made on their drafts. This possibility could have been investigated during the interviews but the interviews appear to have focused on determining whether the L2 writers felt coded or uncoded error feedback was more helpful; the interviews, if conducted in such a way, could have determined if L2 writers considered the task as a revision task or simply a memorization task. Furthermore, the percentage of change in an L2 writer's errors does not tell very much about the accuracy of subsequent writings produced. This instead, may simply be providing a measure of how the L2 writers improved on learning how to complete the task given: being able to memorize which errors were marked by the teacher on a draft and then marking those

same errors on an unmarked copy of the draft. Likewise, the artificiality and lack of authenticity of the task, especially for the control group, could have affected L2 writers' motivation and hence performance. Also, the practice effect has to be taken into account. After the first receptive test, the L2 writers probably learnt that in the subsequent tests they would have to behave in certain ways in order to perform well. Conversely, writers might have become bored (as shown in the interview) and as a result their level of attention or motivation might decline, which could affect the results of the study. This is especially true for the control group. In fact, the receptive test was erroneously used "...to discover the impact of the feedback procedures on learners' ability to recognize and correct errors in their writing work" (p. 498). Given this purpose, a different test should have been designed. In other words, a pre- and post-test that requires writers to correct errors should have been constructed. In addition, upon further inspection, it becomes clear that the experimental and control groups were given a different treatment, where the experimental group was given access to reference resources and peer/teacher assistance, the control group was not. This could have made a direct impact on the writers' performance, as writers in the experimental group were provided with grammar support beyond error feedback but not the control group.

#### **Improvement or misleading metric**

The formula used by Sampson (2012) to calculate the percentage of change in an L2 writer's errors is also questionable. We understand that the purpose of the research was to investigate whether uncoded or coded corrections are more effective, but using the formula suggested by Sampson does not control well for inflation of results. For example, using this formula, it is possible to calculate a smaller percentage in the reduction of errors for a learner that has in fact produced significantly fewer errors and vice versa. Take the following calculations for the fictional Writers A and B as examples: Writer A  $1 - 5/5 * 100 = -80$ ; Writer B  $30 - 75/75 * 100 = -60$ . Writer A produced 1 error on the fourth writing and 5 errors on the first writing. Using the formula suggested by Sampson shows an 80 % decrease in errors. Writer B produced 30 errors on the fourth writing and 75 errors on the first writing. Following the formula shows only a 60 % decrease in errors. In my example, Writer B clearly has progressed more than Writer A but Sampson's formula prevents this from showing. Although Sampson makes note that the learners' percentage change in error frequencies "...showed an alternating pattern of increasing and decreasing success from one test to the next..." (p. 499), we must reiterate our previous concern. Was this due to L2 learners' awareness of correct forms or instead their familiarity of the task and what was expected of them when completing the task? A better picture of L2 writers reduction of errors may have been shown by using a formula or statistical analysis that took into consideration both accurate and inaccurate usage of certain grammatical structures, vocabulary, and punctuation. In addition, no information was provided on how the errors were coded to ensure consistency. It is unclear if a second coder was involved in categorizing errors and if any reliability analysis was performed.

#### **Methodological issues**

Secondly, Sampson (2012) reports that "[a]nything learners had written beyond 150 words was not corrected by the teacher, to ensure equality of sampling across learners'

texts” (pg. 498). From the description of the writing tasks given to learners by the teacher, one can postulate these were narratives. Narrative writing roughly contains three sections: setup, conflict, and resolution (Folse et al. 1999). Although we understand the need to control for sampling across writing, we question whether selecting the first 150 tokens is an adequate strategy. It may be better, for instance, to have taken into consideration both the amount of writing (i.e., 150 tokens) and also the section of the writing. Particular grammar structures or word usage will be found in the setup of narrative discourse that may not appear in other sections (Folse et al. 1999). Since it was not reported whether the first 150 tokens included only the setup or led into other parts of the discourse, it is difficult to determine if this could have affected the results. The wording of the paper further prevents readers from knowing whether all errors produced by the L2 writers were corrected or simply all errors produced in the first 150 tokens of their writings were corrected. Sampson states “...it was necessary to give feedback on all the errors in each piece of work” (p. 498) but previously in the paper had stated “Anything learners had written beyond 150 words was not corrected by the teacher, to ensure equality of sampling across learners’ texts” (p. 498). If in fact only the first 150 tokens were corrected, this was bound to have had an effect on learners’ subsequent writings in that they may have used avoidance as a strategy to decrease the number of error types or may have focused more on the first 150 tokens in their writing while devoting less attention to later parts of their writings (Truscott 2007). A more viable alternative would be to first calculate error gravity by dividing the number of errors by the total number of tokens in a student’s writing and then working out an error ratio for each error type (see Kao and Wible 2014). Such an approach would enable a researcher to determine if coded error feedback may be particularly useful for specific error types.

### Opaque coding

Lastly, the correction symbols from Olsher (1995) adopted by Sampson (2012) were provided but examples were only provided of how uncoded errors were marked. Since Sampson claims the usage of symbols are helpful to L2 writers, then examples of their use with the L2 writer data collected should have been provided. It is indeterminable whether any systematic method of placement of the symbols was implemented since this was not reported. Furthermore, although underlining of errors was considered as uncoded error feedback, the examples of uncoded feedback given by Sampson are problematic. Specifically, underlining a single letter as in the example “My birthday is in januray.” is likely to signal a different type of result from L2 writers than for underlining of a single word as in the example “We luve chocolate.” or the underlining of multiple words in the example “I you see will later.” The different types of underlining in themselves could be considered a type of coded feedback. Besides, in the example given by Sampson for the error type “Add something”, it is unclear how to underline an error of something that is missing. In the example provided “It is \_\_\_\_ beautiful afternoon” a space exists, but in an L2 writer’s draft this space will not exist since the writer will have left something out. So where should the underlining appear? This is another reason why details regarding the placement of both uncoded and coded feedback should have been provided. Furthermore, the concept of “error” is not explained well in the study. Specifically, “gently person” is marked as an error but could be labeled as

either “word formation,” “spelling,” or “wrong word.” The term “add something” seems too general and would prevent the L2 writer from addressing the error. “Reverse word order” and “word order mistake” seem too similar to constitute two separate categories. “Rewrite” is a label for meaning-oriented errors or content errors, which are not easily fixed through a single revision by an L2 writer.

## Summary

ESL and EFL writing instructors are bound to provide some types of written feedback on L2 writers’ drafts. This is often done with the explicitness of corrections in previous research (i.e. whether corrections should be coded or uncoded). However, there is a serious mismatch between research findings and classroom applications in the area of error correction (Lee 2013). What we are arguing in this commentary is that Sampson (2012) might overestimate the benefit of teachers’ feedback on students’ language errors. The effectiveness of feedback might be attributed to the discussed or other extraneous variables. In addition, Sampson fails to provide an explicit articulation of how corrections are offered and how specific errors are corrected. Since a good error correction research design should be able to be duplicated (Ferris 2004; Guénette 2007), Sampson’s study seems to have a problem in terms of this point particularly in one of the crucial aspects of error correction studies - “replicability”. We therefore suggest future researchers should value the aspect of replicability to bridge the gap between research findings and classroom applications. We anticipate that our research colleagues in the field of L2 error correction will take up this quest and encourage anyone wishing to work jointly on such a project to initiate collaboration.

## Authors’ contributions

BLR approached CWK with the prospect of writing the commentary and discussed key points of the commentary. CWK conducted the literature review. BLR wrote the critique and CWK incorporated the literature into the critique. BLR added additional content and revised the submission. CWK edited the manuscript before submission. BLR and CWK discussed reviewer comments and then BLR edited the manuscript before resubmission. Both authors read and approved the final manuscript.

## Authors’ information

Barry Lee Reynolds, Ph.D. is Assistant Professor of English Education in the Faculty of Education at the University of Macau, Macau SAR, China. His research interests include L1/L2 incidental vocabulary acquisition, second language vocabulary learning, L2 Reading and Writing Instruction, and other areas of Applied Linguistics. His published work appears in *TESOL Quarterly*, *Reading Research Quarterly*, *Applied Linguistics Review*, *English Today*, *Computers & Education*, *British Journal of Educational Technology*, among others.

Dr. Chian-Wen Kao is currently a research fellow in the Center for Teaching and Learning at National Taipei University of Business, where he has also taught as academic faculty in the Department of Applied Foreign Languages. His research interests include second language writing developments, digital game based learning and second language learning strategies. His recent publications appear in *Journal of Chinese Language Teaching*, *Education Journal*, *English Teaching & Learning* and *TESL-EJ: Teaching English as a Second or Foreign Language*, among others.

## Competing interests

We have read and understood the *Asian-Pacific Journal of Second and Foreign Language Education* policy on declaration of interests and declare that we have no financial, professional or personal competing interests that might have influenced the performance or presentation of the work described in this manuscript.

## Author details

<sup>1</sup>Faculty of Education, University of Macau, Room 1017, FED Building (E33), Avenida da Universidade, Taipa, Macau SAR.

<sup>2</sup>Section of Language Learning, Center for Teaching and Learning, National Taipei University of Business, No. 321, Section 1, Jinan Rd, Zhongzheng District, Taipei City 100, Taiwan ROC.

Received: 30 August 2016 Accepted: 18 October 2016

Published online: 25 October 2016

## References

Ferris, D. (2004). The “grammar correction” debate in L2 writing: where are we, and where do we go from here? (and what do we do in the mean time...?). *Journal of Second Language Writing*, 13, 49–62.

- Folse, K. S., Muchmore-Vokoun, A., & Solomon, E. V. (1999). *Great essays*. Boston: Houghton Mifflin Company.
- Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16, 40–53.
- Kao, C.-W., & Wible, D. (2014). A meta-analysis on the effectiveness of grammar correction in second language writing. *English Teaching & Learning*, 38(3), 29–69. doi:10.6330/ETL.2014.38.3.02.
- Lee, I. (2013). Research into practice: written corrective feedback. *Language Teaching*, 46, 108–119.
- Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. New York: Routledge.
- Olsher, D. (1995). *Words in motion: an interactive approach to writing*. Oxford: OUP.
- Sampson, A. (2012). Coded and uncoded error feedback: effects on error frequencies in adult Colombian EFL learners' writing. *System*, 40(4), 494–504. <http://dx.doi.org/10.1016/j.system.2012.10.001>.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16, 255–272.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---